

Approximate Bayesian Computation and Model Assessment for Repulsive Spatial Point Processes

Shinichiro Shirota*

Department of Statistical Science, Duke University, U.S.

and

Alan. E. Gelfand†

Department of Statistical Science, Duke University, U.S.

August 29, 2016

Abstract

In many applications involving spatial point patterns, we find evidence of inhibition or repulsion. The most commonly used class of models for such settings are the Gibbs point processes. A recent alternative, at least to the statistical community, is the determinantal point process. Here, we examine model fitting and inference for both of these classes of processes in a Bayesian framework. While usual MCMC model fitting can be available, the algorithms are complex and are not always well behaved. We propose using approximate Bayesian computation (ABC) for such fitting. This approach becomes attractive because, though likelihoods are very challenging to work with for these processes, generation of realizations given parameter values is relatively straightforward. As a result, the ABC fitting approach is well-suited for these models. In addition, such simulation makes them well-suited for posterior predictive inference as well as for model assessment. We provide details for all of the above along with some simulation investigation and an illustrative analysis of a point pattern of tree data exhibiting repulsion. R-code and datasets are included in the supplementary material.

Keywords: determinantal point processes; Gibbs point processes; model checking; summary statistics

*PhD student, Department of Statistical Science, Duke University, U.S. (E-mail: ss571@stat.duke.edu)

†Professor, Department of Statistical Science, Duke University, U.S. (E-mail: alan@stat.duke.edu)

1 Introduction

There is increasing interest in analyzing spatial point process data. In the literature, the most widely adopted class of models are nonhomogeneous Poisson processes (NHPP) or, more generally, Cox processes, including log Gaussian Cox processes (LGCP) (see Møller and Waagepetersen (2004) and references therein). Such models assume conditionally independent event locations given the process intensity. However, in many applications, we find evidence of clustering or of inhibition. Here, we focus on inhibition and refer to associated models as repulsive spatial point processes. Most common in this setting are the Gibbs point processes (here, denoted as GPP) (see, e.g., Illian et al. (2008)). These processes specify the joint location density, up to normalizing constant, in the form of a Gibbs distribution, introducing potentials on cliques of order 1 but also potentials on cliques of higher order, which capture interaction. The most familiar example in the literature is the Strauss process and its extreme version, the hardcore process (Strauss (1975) and Kelly and Ripley (1976)). An attractive alternative is the determinantal point process (here, denoted as DPP). Though these processes have some history in the mathematics and physics communities, they have only recently come to the attention of the statistical community thanks, most notably, to recent efforts of Jesper Møller and colleagues. See, for instance, Lavancier et al. (2015).

The contribution of this paper is to explore both the GPP and DPP models with regard to application. We will briefly summarize and compare their respective properties. We will consider model fitting in a Bayesian framework. Markov chain Monte Carlo (MCMC) model fitting has been proposed for both GPPs and DPPs (Møller et al. (2006), Affandi et al. (2013), Affandi et al. (2014), Goldstein et al. (2015)). The algorithms are quite complex and implementation can often result in poorly behaved chains with concerns regarding posterior convergence. Here, we propose much simpler model fitting using approximate Bayesian computation (ABC). ABC is particularly promising for GPPs and DPPs since these processes allow straightforward simulation of point pattern realizations given parameter values. Additionally, such simulation facilitates posterior inference as well as consideration of model adequacy and model comparison.

ABC methods are now attracting considerable attention (Pritchard et al. (1999), Beau-

mont et al. (2002), Marjoram et al. (2003), Sisson and Fan (2011) and Marin et al. (2012)). The scope of ABC applications is also increasing, e.g., population genetics (Beaumont et al. (2002)), multidimensional stochastic differential equations (Picchini (2013)), macroparasite populations (Drovandi and Pettitt (2011)) and the evolution of the HIV-AIDS (Blum and Tran (2010)). As for spatial statistical application, Erhardt and Smith (2012) implemented ABC for max-stable processes in order to model spatial extremes and Soubeyrand et al. (2013) applied ABC with functional summary statistics to fit a cluster and a marked spatial point process.

We briefly review some of the repulsive point process literature. The Gibbs point process offers a mechanistic model with interpretable parameters and has been used for modeling repulsive point patterns in environmental science and biology (Stoyan and Penttinen (2000), Mattfeldt et al. (2007), King et al. (2012) and Goldstein et al. (2015)). The main challenge for fitting models using these processes is that likelihoods have intractable normalizing constants which depend on parameters. Hence, likelihood inference is difficult (Møller and Waagepetersen (2004)) and standard Bayesian inference using Markov chain Monte Carlo (MCMC) is not directly available. Maximum pseudo-likelihood estimation was proposed in Besag (1975), Besag (1986) and Jensen and Møller (1991). These estimators show poor performance in the presence of strong inhibition (e.g., Huang and Ogata (1999)). In the Bayesian framework, a clever auxiliary variable MCMC strategy has been developed by Møller et al. (2006) and extended by Murray et al. (2006). However, perfect simulation within the MCMC algorithm is needed along with approximations.

Determinantal point processes arise in random matrix theory and quantum physics, and are investigated mainly by probabilists with a machine learning perspective (see, Macchi (1975), Hough et al. (2006) and Kulesza and Taskar (2012)). In contrast to GPPs, DPPs have analytical expressions for moments and approximate likelihoods. In particular, Lavancier et al. (2015) investigated statistical inference for and properties of spatial DPPs along with approximation for evaluation of the likelihood. The pattern of repulsiveness exhibited by DPPs may be different from that of GPPs so, making a modeling choice for a spatial repulsive point process may be unclear.

We propose to implement the ABC algorithm based on ABC-MCMC proposed by Mar-

joram et al. (2003) for fitting both GPPs and DPPs. We include discussion about how to choose summary statistics, kernel functions and, tuning user specific parameters. Again, the attractiveness of ABC for repulsive point processes rests in the fact that it is straightforward to generate realizations under these point processes given parameter values. This enables the ABC presumption: randomly draw parameters and then randomly draw point patterns given parameters. Further, with posterior inference achieved for the model parameters, we can use composition sampling to draw posterior predictive point patterns enabling posterior inference about features of point patterns realized under the models. In addition, through these posterior samples of point patterns, we can propose model assessment for repulsive point processes, following the discussion in Leininger and Gelfand (2016).

We offer simulation investigation of the model fitting approach. We also analyze a real dataset consisting of 89 trees from the Blackwood region in Duke Forest in the US. All of the statistical analyses presented here were implemented with R. In particular, the `spatstat` package (Baddeley and Turner (2005)) is utilized for simulating point patterns and calculating functional summary statistics associated with DPPs and GPPs. The R code is attached as supplementary material.

The format of the paper is as follows. Section 2 introduces the repulsive point processes we investigate and their associated inference. Section 3 provides the development of our ABC fitting strategy for repulsive point processes along with model validation for and comparison strategy between different repulsive point processes. In Section 4, we investigate the proposed algorithms with some simulation studies while in Section 5 we analyze the foregoing forest data. Finally, Section 6 offers discussion about the pros and cons of our approach as well as potential future work.

2 Repulsive Spatial Point Processes

Here, we provide a brief review of the two classes of repulsive point processes that we consider. GPPs are taken up in Section 2.1, DPPs in Section 2.2. We consider a bounded spatial domain $D \subset \mathbb{R}^2$ and denote the (finite) point pattern over D by $\mathbf{x} = \{x_1, \dots, x_n\}$.

2.1 Gibbs Point Processes

The joint location density for a Gibbs process takes the form of a Gibbs distribution (e.g., Georgii (1988)). In particular, we say that a point process model is a finite Gibbs process if, for n locations, its location density is

$$\pi(\mathbf{x}) = \exp(-Q(\mathbf{x})) \quad (1)$$

with regard to a homogeneous Poisson process (HPP) with unit intensity (e.g., van Lieshout (2000) or Møller and Waagepetersen (2004)). In general,

$$Q(x_1, x_2, \dots, x_n) = c_0 + \sum_{i=1}^n h_1(x_i) + \sum_{i \neq j} h_2(x_i, x_j) + \dots + h_n(x_1, x_2, \dots, x_n). \quad (2)$$

In (2), the h 's are potentials of order 1, 2, ..., n , respectively, each symmetric in its arguments. Here, c_0 is a *normalizing* constant to make $\pi(\mathbf{x})$ integrate to 1 over D^n . Denoting parameters in the Gibbs potentials by $\boldsymbol{\theta}$, the normalizing constant becomes $c_0(\boldsymbol{\theta})$. Evidently it can not be calculated analytically and, in fact, is computationally intractable.

With potentials only of order 1, we obtain a nonhomogeneous Poisson process (NHPP) with $\lambda(x) \propto e^{-h_1(x)}$. Higher order potentials capture/control interaction; customarily only h_1 and h_2 are included. To guarantee integrability, we must take $h_2 \geq 0$ which implies that we can only capture inhibition. In other words, if we require $Q(x_1, x_2) \geq c_0 + h_1(x_1) + h_1(x_2)$, this means for pairs of points at a given distance, $\pi(x_1, x_2)$ puts less mass under the Gibbs specification than with the corresponding NHPP; we encourage inhibition. If $h_1(x)$ is constant, we have a homogeneous Gibbs process. The most common form for h_2 is $\phi(\|x - x'\|)$, e.g., $\phi(\|x - x'\|) = e^{-\|x - x'\|^2/\eta^2}$. The Papangelou conditional intensity (Illian et al. (2008)) becomes

$$\lambda(x|\mathbf{x}) = \exp(-(h_1(x) + \sum_{i=1}^n \phi(\|x - x_i\|))). \quad (3)$$

Attractively, the unknown normalizing constant cancels in the conditional intensity.

The Strauss process is a GPP with density often written as (e.g., Møller and Waagepetersen (2004))

$$\pi(\mathbf{x}) = \beta^{n(\mathbf{x})} \gamma^{s_R(\mathbf{x})} / c(\beta, \gamma), \quad \mathbf{x} \in D^n \quad (4)$$

where $\beta > 0$, $0 \leq \gamma \leq 1$, $n(\mathbf{x})$ is the number of points, $c(\beta, \gamma)$ is the normalizing constant and

$$s_R(\mathbf{x}) = \sum_{\{\xi, \eta\} \subseteq \mathbf{x}} \mathbf{1}(\|\xi - \eta\| \leq R), \quad \xi, \eta \in D \quad (5)$$

is the number of R -close pairs of points in \mathbf{x} . Given R , $n(\mathbf{x})$ and $s_R(\mathbf{x})$ are sufficient statistics for (β, γ) . γ is an interaction parameter indicating the degree of repulsion. Large value of γ suggest weak repulsion while small values of γ indicate strong repulsion. $\gamma = 0$ provides the hardcore Strauss process which does not allow occurrence of any points within radius R . $\gamma = 1$ provides a homogeneous Poisson process.

For GPPs, since $c_0(\boldsymbol{\theta})$ cancels out of the Papangelou conditional intensity, the pseudo-likelihood, in log form, $\log PL(\mathbf{x}|\boldsymbol{\theta}) = -\int_D \lambda(u|\mathbf{x}, \boldsymbol{\theta}) du + \sum_i \log \lambda(x_i|\mathbf{x}, \boldsymbol{\theta})$, has been proposed (Besag (1975)) yielding the maximum pseudo-likelihood estimator. Although the maximum pseudo-likelihood estimator is consistent (see, Jensen and Møller (1991) and Mase (1995)), the performance of the maximum pseudo-likelihood estimator is poor in the case of a small number of points and strong repulsion (Huang and Ogata (1999)).

The pseudo-likelihood can be used for MCMC in the Bayesian framework (e.g. King et al. (2012)). Møller et al. (2006) and Berthelsen and Møller (2008) proposed a clever auxiliary variable MCMC method (AVM) where, conveniently, $c_0(\boldsymbol{\theta})$ cancels out of the Hastings ratio. The challenge is to obtain the conditional density of the auxiliary variable. A partially ordered Markov model is used to approximate this density. A similar approach is the exchange algorithm proposed by Murray et al. (2006). Both algorithms require perfect simulation from the likelihood given $\boldsymbol{\theta}$ for each MCMC iteration. Although, perfect simulation is available for GPPs, this step can be computationally burdensome and obtaining a good acceptance rate is difficult.

More recently, Liang (2009) proposed the double MCMC algorithm which does not require perfect simulation from the likelihood. It only requires simulation from the Markov transition kernel and is faster than the AVM and exchange algorithms but convergence to the stationary distribution is not guaranteed. Goldstein et al. (2015) implement this algorithm, with an application, for a class of GPP models.

For ABC, we need simulation of realizations of a GPP given parameter values. This is usually based on a birth-and-death algorithm (e.g., Geyer and Møller (1994), Møller and

Waagepetersen (2004) and Illian et al. (2008)). The simulation algorithm we use to generate the point pattern is "dominated coupling from the past" (Kendall and Møller (2000)) as implemented by Berthelsen and Møller (2002) and Berthelsen and Møller (2003). This algorithm can be called as a default setting in `spatstat` (Baddeley and Turner (2005)).

2.2 Determinantal Point Processes

Theoretical properties for DPPs are investigated by Hough et al. (2006) and elaborated from a statistical perspective by Lavancier et al. (2015). The n -th order product density function (Møller and Waagepetersen (2004)) for DPPs arises as determinants of covariance kernels. We briefly review the definition and theoretical properties of DPPs drawn from Lavancier et al. (2015).

Suppose, for a finite spatial point process \mathbf{x} on a compact set $D \subset \mathbb{R}^2$,

$$\tau^{(n)}(x_1, \dots, x_n) = \det\{[C](x_1, \dots, x_n)\}, \quad (x_1, \dots, x_n) \in D^n, \quad n = 1, 2, \dots \quad (6)$$

is the n -th order product density function. Here, $C(x_i, x_j)$ is a covariance kernel for locations x_i and x_j and $\det\{[C](x_1, \dots, x_n)\}$ denotes the determinant with (i, j) -th entry, $C(x_i, x_j)$. Then \mathbf{x} is called a *DPP* with kernel C restricted to a compact set D , and we write $\mathbf{x} \sim DPP_D(C)$. Hence, the first order density function is $\tau(x) = C(x, x)$ and the pair correlation function is

$$g(x_1, x_2) = 1 - \frac{C(x_1, x_2)C(x_2, x_1)}{C(x_1, x_1)C(x_2, x_2)}, \quad \text{if } C(x_1, x_1) > 0 \quad \text{and} \quad C(x_2, x_2) > 0 \quad (7)$$

whereas it is 0 otherwise. Since C is a covariance kernel, then $\tau^{(n)}(x_1, \dots, x_n) \leq \tau(x_1) \dots \tau(x_n)$ for any $n > 1$, implying repulsion, and $g \leq 1$ (Lavancier et al. (2015)).

For a given covariance function, existence conditions for the DPP are supplied in Lavancier et al. (2015). Here, we confine ourselves to three real valued covariance functions over $d = 2$ dimensional space: (1) the Gaussian kernel, $C_G(x, y) = \tau \exp(-\|x - y\|^2/\sigma^2)$, (2) the Whittle-Matérn, $C_{WM}(x, y) = \tau \frac{2^{1-\nu}}{\Gamma(\nu)} (\|x - y\|/\sigma)^\nu K_\nu(\|x - y\|/\sigma)$, $\nu > 0$, and (3) the Generalized Cauchy, $C_C(x, y) = \tau(1 + \|x - y\|^2/\sigma^2)^{-\nu-1}$, $\nu > 0$, where τ is the variance parameter. For all three kernels, with $d = 2$, the existence of the process is guaranteed

when $\sigma \leq \sigma_{max} = (U/\tau)^{1/2}$, where $U = \pi^{-1}$ for the Gaussian, $U = \Gamma(\nu)/(4\pi\Gamma(\nu + 1))$ for the Whittle-Matérn and $U = \Gamma(\nu + 1)/(\pi\Gamma(\nu)) = \nu/\pi$ for the Cauchy.

The Gaussian kernel is asserted to provide *moderate* repulsion. A specification enabling *stronger* repulsion can be obtained through the spectral density. Lavancier et al. (2015) supply the power exponential DPP model induced by the spectral density,

$$\varphi(x) = \tau \frac{\alpha^2}{\pi\Gamma(2/\nu + 1)} \exp(-\|\alpha x\|^\nu), \quad \tau \geq 0, \quad \alpha > 0, \quad \nu > 0 \quad (8)$$

For fixed τ and ν , existence is guaranteed when $\alpha \leq \alpha_{max} = \sqrt{\Gamma(2/\nu + 1)\pi/\tau}$. $\alpha \approx \alpha_{max}$ shows strong repulsiveness. In our simulation examples below, we illustrate the use of the DPP with both a Gaussian kernel (DDP-G) and a power exponential (DDP-PE) to enable investigation of both moderate and stronger repulsiveness.

Simulation algorithms for DPPs are proposed in e.g., Hough et al. (2006), Scardicchio et al. (2009) and Lavancier et al. (2015). We employ the algorithm in Theorem 2 of Lavancier et al. (2015) which is based on the spectral decomposition of the associated kernel covariance function. Recall the spectral representation of the covariance function, $C(x_i, x_j) = \sum_{k=1}^{\infty} \lambda_k \psi_k(x_i) \bar{\psi}_k(x_j)$ where $(x_i, x_j) \in D \times D$. The existence of the DPP is guaranteed if all of the eigenvalues, $\lambda_k \leq 1$. Define the *projection* kernel function as $K(x_i, x_j) = \sum_{k=1}^{\infty} G_k \psi_k(x_i) \bar{\psi}_k(x_j)$. Here, the G_k are independent Bernoulli variables with mean λ_k for $k = 1, 2, \dots$. The DPP with this projection kernel has the same distribution as the $DPP_D(C)$.

Exact simulation of a determinantal point process model involves an infinite series, which has no analytical solution except for a few kernel choices (e.g., Macchi (1975)), which might be insufficient to describe the interaction structure in real datasets (Lavancier et al. (2015)). So, we need a truncation of the infinite sum. This truncation is implemented by $(E[\tilde{N}(D)] - E[N(D)]) < 0.01E[N(D)]$ where $N(D) \sim \sum_{k=1}^{\infty} G_k$ (so $E[N(D)] = \sum_{k=1}^{\infty} \lambda_k$) and the truncation, $\tilde{N}(D)$, is generated by approximate simulation (Lavancier et al. (2015) and Baddeley and Turner (2005)). With $\mathbf{v}(x) = (\psi_1(x), \dots, \psi_n(x))^T$, the simulation algorithm given by Lavancier et al. (2015) is

1. Set $n = \sum_{k=1}^{\infty} G_k$

2. Sample x_n from the distribution with density $p_n(x) = \|\mathbf{v}(x)\|^2/n$, $x \in D$ and set $\mathbf{e}_1 = \mathbf{v}(x_n)/\|\mathbf{v}(x_n)\|$

3. For $i = n - 1$ to $i = 1$, sample x_i from the distribution with density

$$p_i(x) = \frac{1}{i} \left\{ \|\mathbf{v}(x)\|^2 - \sum_{j=1}^{n-i} |\mathbf{e}_j^* \mathbf{v}(x)|^2 \right\}, \quad x \in D.$$

Set $\mathbf{w}_i = \mathbf{v}(x_i) - \sum_{j=1}^{n-i} \{\mathbf{e}_j^* \mathbf{v}(x_i)\} \mathbf{e}_j$, $\mathbf{e}_{n-i+1} = \mathbf{w}_i/\|\mathbf{w}_i\|$.

4. Return $\{x_1, \dots, x_n\}$

Rejection sampling is used in Step 3 with a uniform density λ_0 over D and acceptance probability given by $p_i(x)/\lambda_0$, where λ_0 is an upper bound on $p_i(x)$ for $x \in D$.

Turning to inference, continuous DPPs (DPPs defined on continuous space $D \in \mathbb{R}^d$), e.g., the spatial DPP of interest to us, typically require the infinite dimensional spectral decomposition, which are not analytically available except for a few kernel choices (e.g., Macchi (1975)), to evaluate the likelihood and to do exact simulation. The slice sampling approach by Affandi et al. (2014), calculating upper and lower bounds of DPP likelihoods, can be applied but requires low rank approximation of the spectral representation (Affandi et al. (2013)).

For stationary continuous DPPs, the likelihood function for $\mathbf{x} = \{x_1, \dots, x_n\} \in D^n$ is defined as

$$f(\mathbf{x}) = \exp(|D| - H) \det\{\tilde{C}(x_1, \dots, x_n)\} \quad (9)$$

where $H = -\log[P(N(D) = 0)] = -\sum_{k=1}^{\infty} \log(1 - \lambda_k)$ and $\tilde{C}(x, y) = \sum_{k=1}^{\infty} \tilde{\lambda}_k \psi_k(x) \bar{\psi}_k(y)^T$, $x, y \in D$, $\tilde{\lambda}_k = \lambda_k/(1 - \lambda_k)$. Since the likelihood involves an infinite dimensional spectral decomposition, Lavancier et al. (2015) consider the maximum likelihood estimator based on an approximate likelihood constructed on a rectangular region by using a Fourier basis (see Lavancier et al. (2015) for details). For a large number of points, calculation of components of the covariance kernel, i.e., $C(x_i, x_j)$ for each pair of (x_i, x_j) , is computationally costly, even using the Fast Fourier transformation. There is also potential sensitivity to the resolution of the grid. For parametric families of DPPs, the Papangelou conditional intensity is not easier to calculate than the likelihood itself, so pseudo-likelihood estimates

are not easily available as with GPPs. Bayesian inference using MCMC will be even more challenging.

Lastly, as Lavancier et al. (2015) note, handling of the DPP for a non-rectangular region is not clear. Embedding a non-rectangular observation window W in a rectangular region R , yielding a missing data problem, is one possible approach.

3 Approximate Bayesian Computation for Repulsive Point Processes

Let \mathbf{y} be the observed point pattern and \mathbf{x} be a simulated point pattern. For a Bayesian model of the form $\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$, ABC consists of three steps: (1) generate $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$, (2) generate $\mathbf{x} \sim \pi(\mathbf{x}|\boldsymbol{\theta})$, (3) compare summary statistics for the generated \mathbf{x} , $\mathbf{T}(\mathbf{x})$, with those of the observed data, $\mathbf{T}(\mathbf{y})$, and accept $\boldsymbol{\theta}$ if $\Psi(\mathbf{T}(\mathbf{x}), \mathbf{T}(\mathbf{y})) < \epsilon$ for a selected *kernel(distance)* measure Ψ . Accepted $\boldsymbol{\theta}$ are samples from the approximate posterior distribution, $\pi_\epsilon(\boldsymbol{\theta}|\mathbf{T}(\mathbf{y}))$. Approximation error relative to the exact posterior distribution $\pi(\boldsymbol{\theta}|\mathbf{y})$ comes from the choice of $\mathbf{T}(\cdot)$, Ψ , and ϵ . If $\mathbf{T}(\cdot)$ is a sufficient statistic for $\boldsymbol{\theta}$, then $\pi(\boldsymbol{\theta}|\mathbf{T}(\mathbf{y})) = \pi(\boldsymbol{\theta}|\mathbf{y})$ and, given Ψ , the only approximation error is from ϵ . Since sufficient statistics are not usually available, the selection of informative summary statistics $\mathbf{T}(\cdot)$ is critically important. Small values of ϵ are desired but require more simulation of $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$ and $\mathbf{x} \sim \pi(\mathbf{x}|\boldsymbol{\theta})$. Again, with regard to simulation of \mathbf{x} : (i) for the GPP, we can utilize perfect simulation, (ii) for the DPP, we can use the Fourier basis approximate simulation method by truncating infinite sums for DPPs whose kernels don't have an analytical spectral representation.

3.1 Summary Statistics

Second order summary statistics play a fundamental role in the statistical analysis of spatial point patterns because they illuminate clustering or inhibition behavior. The most common choice is Ripley's K -function (Ripley (1976) and Ripley (1977)). For a stationary point process, the K -function with radius r , K_r , is the *expected* number of the remaining points

in the pattern within distance r from a typical point. The empirical estimator of K_r is

$$\hat{K}_r(\mathbf{x}) = |D| \sum_{\xi, \eta \in \mathbf{x}: \xi \neq \eta} \frac{\mathbf{1}[0 < \|\xi - \eta\| \leq r]}{n(n-1)} e(\xi, \eta) \quad (10)$$

where $e(\cdot, \cdot)$ is an edge correction factor (e.g., Illian et al. (2008)). The variance stabilized version, the L -function (Besag (1977)), $\hat{L}_r(\mathbf{x}) = \sqrt{\hat{K}_r(\mathbf{x})/\pi}$ is often preferred. Under the Poisson process model, the expected value of $\hat{L}_r(\mathbf{x})$ is r . When $\hat{L}_r(\mathbf{x}) - r < 0$ for small to modest values of $r > 0$, the point pattern suggests an inhibitive point process model (see e.g. Møller and Waagepetersen (2004)).

With the Strauss GPP, given the interaction radius R , from (5), $s_R(\mathbf{x})$, is a sufficient statistic and hence, the appropriate summary statistics would be $\mathbf{T} = (\log n(\mathbf{x}), K_R(\mathbf{x}))$. In practice, R is not known but we can choose a radius R though profile pseudo likelihoods. Alternatively, creating a set of R values yields a set of summary statistics.

For DPPs, repulsiveness of DPPs is determined by the covariance kernel function; there is no notion of a radius. Nonetheless, we propose to use a set $\{\hat{K}_r(\mathbf{x})\}$ evaluated at M selected values of r over the range $[0, r_{max}]$. That is, with $r_0 = 0, r_1, \dots, r_M = r_{max}$, calculate $\hat{K}_{r_1}(\mathbf{x}), \dots, \hat{K}_{r_M}(\mathbf{x})$. Sensitivity to the number of and choice of R 's is considered below. So, analogous to the GPP, we assume $\mathbf{T} = (T_1, \mathbf{T}_2)$, where $T_1 = \log n(\mathbf{x})$ and the components of \mathbf{T}_2 are $T_{2,r}(\mathbf{x}) = \sqrt{\hat{K}_r(\mathbf{x})}$.

Soubeyrand et al. (2013) propose an optimized weight function ABC strategy for functional summary statistics with application to spatial point processes (the Neyman-Scott process and a marked point process). For smaller M (≤ 20), they calculate an optimized weighted *distance* by the Nelder-Mead algorithm between the simulated and observed functional statistics by minimizing the mean square error of a point estimate of $\boldsymbol{\theta}$. For larger M , the Nelder-Mead algorithm is not available. So, Soubeyrand et al. (2013) adopt lower dimensional piecewise constant weight parameters which can be optimized through the algorithm to obtain manageable computation time.

3.2 Explicit specification of our ABC algorithm

The ABC algorithm we adopt is based on a semi-automatic approach proposed by Fearnhead and Prangle (2012). They argue that the optimal choice of $\mathbf{T}(\mathbf{y})$ is $E(\boldsymbol{\theta}|\mathbf{y})$ and

then discuss how to construct $E(\boldsymbol{\theta}|\mathbf{y})$. They consider a linear regression approach to construct the summary statistics through a pilot run. In our setting, we generate L sets of $\{\boldsymbol{\theta}_\ell, \mathbf{x}_\ell\}_{\ell=1}^L$ (choice for L discussed below). Then, we implement a linear regression for $E(\boldsymbol{\theta}_\ell|\mathbf{y}) = \mathbf{a} + \mathbf{b}\boldsymbol{\eta}(\mathbf{x}_\ell, \mathbf{y})$ where $\boldsymbol{\eta}(\mathbf{x}_\ell, \mathbf{y})$ is a vector of functions of the summary statistics constructed from the simulated and observed point patterns ¹. We take $\boldsymbol{\eta}(\mathbf{x}, \mathbf{y}) = (\eta_1(\mathbf{x}, \mathbf{y}), \eta_2(\mathbf{x}, \mathbf{y}))$ with, for $r = 1, \dots, M$,

$$\eta_1(\mathbf{x}, \mathbf{y}) = \log n(\mathbf{x}) - \log n(\mathbf{y}), \quad \eta_{2,r}(\mathbf{x}, \mathbf{y}) = \left| \sqrt{\hat{K}_r(\mathbf{x})} - \sqrt{\hat{K}_r(\mathbf{y})} \right|^2 \quad (11)$$

After obtaining $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ by least squares, we can calculate $\hat{\boldsymbol{\theta}}_* = \hat{\mathbf{a}} + \hat{\mathbf{b}}\boldsymbol{\eta}(\mathbf{x}^*, \mathbf{y})$ for any simulated \mathbf{x}^* . We set $\hat{\boldsymbol{\theta}}_{obs} = \hat{\mathbf{a}}$ and specify our distance function for the ABC through $\Psi(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\theta}}_{obs})$ with Ψ specified below. To facilitate the regression, we take a log transformation of the parameter vector, e.g., $\boldsymbol{\theta} = (\log \beta, \log \gamma)$ for the Strauss process. Given the results of the pilot run, the approach proposed of Fearnhead and Prangle (2012) implements the ABC-MCMC algorithm by Marjoram et al. (2003). ABC-MCMC is a straightforward extension of the standard MCMC framework to ABC; convergence to the approximate posterior distribution, $\pi_\epsilon(\boldsymbol{\theta}|\mathbf{T}(\mathbf{y}))$, is guaranteed. Specifically, with t denoting iterations and $q(\cdot|\cdot)$ denoting a proposal density,

1. Let $t = 1$.
2. Generate $\boldsymbol{\theta}^* \sim q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t-1)})$ and $\mathbf{x}^* \sim \pi(\mathbf{x}|\boldsymbol{\theta}^*)$ and calculate $\hat{\boldsymbol{\theta}}^* = \hat{\mathbf{a}} + \hat{\mathbf{b}}\boldsymbol{\eta}(\mathbf{x}^*, \mathbf{y})$. Repeat this step until $\Psi(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\theta}}_{obs}) < \epsilon$ where $\hat{\boldsymbol{\theta}}_{obs} = \hat{\mathbf{a}}$ and $\Psi(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\theta}}_{obs})$ is defined below.
3. Calculate the acceptance ratio $\alpha = \min \left\{ 1, \frac{\pi(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^{(t-1)}|\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^{(t-1)})q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t-1)})} \right\}$. If $u < \alpha$ where $u \sim \mathcal{U}(0, 1)$ retain $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^*$, otherwise $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$. Return to step 2 and $t \rightarrow t + 1$.

As a distance measure, we use the componentwise sum of quadratic loss for the log of the parameter vector, i.e., $\Psi(\hat{\boldsymbol{\theta}}_\ell, \hat{\boldsymbol{\theta}}_{obs}) = \sum_j (\hat{\theta}_{\ell,j} - \hat{\theta}_{obs,j})^2 / \text{var}(\hat{\theta}_j)$ where $\text{var}(\hat{\theta}_j)$ is the sample variance of j -th component of $\hat{\boldsymbol{\theta}}$. To choose an acceptance rate ϵ , through the pilot run we obtain the empirical percentiles of $\{\Psi(\hat{\boldsymbol{\theta}}_\ell, \hat{\boldsymbol{\theta}}_{obs})\}_{\ell=1}^L$ and then select ϵ according to these

¹Fearnhead and Prangle (2012)) implement linear regression for each component of $\boldsymbol{\theta}$. Since we have a small number of parameters, we keep the notation as linear regression for multivariate responses.

percentiles. Step 2 is the most computationally demanding. We need to simulate the proposed point pattern $\mathbf{x}^* \sim \pi(\mathbf{x}|\boldsymbol{\theta}^*)$ until $\Psi(\hat{\boldsymbol{\theta}}^*, \hat{\boldsymbol{\theta}}_{obs}) < \epsilon$.

With regard to choice of number of summary statistics, when M is large more information is obtained but, if too large, overfitting, relative to the number of points in the point pattern, results. As a strategy for this selection, we specify M with equally spaced r 's, and implement a lasso (Tibshirani (1996)). We determine the penalty parameter for the lasso by cross-validation, and preserve the regression coefficients corresponding to the optimal penalty by using `glmnet` (Friedman et al. (2010)). In our simulation study, we examine sensitivity to the selection of M .

We remark that, in frequentist analysis, the minimum contrast estimator is often used. This is the value of $\boldsymbol{\theta}$ which minimizes $\sum_r |\hat{K}_r(\mathbf{x})^a - K_{\boldsymbol{\theta}}(r)^a|^b$ where $K_{\boldsymbol{\theta}}$ is the theoretical K -function, $\hat{K}_r(\mathbf{x})$ is the empirical estimator for the K -function, and a, b are user-specified parameters (Diggle (2003)). The minimum contrast estimator requires analytical forms for the functional statistics which are not necessarily available for repulsive point processes, e.g., the K function for the Matérn-Whittle kernel function requires numerical methods (though the pair correlation function, whose analytical form is available for the Matérn-Whittle kernel function, is an alternative functional summary statistics for the minimum contrast estimator). ABC does not require analytical expressions for the functional statistics because the approach compares the "estimated" K -function for observed and simulated point patterns. However, if analytical forms for the functional summary statistics are available, the minimum contrast estimator or composite likelihood estimators (Baddeley and Turner (2000) and Guan (2006)) would be available and easy to implement. Furthermore, software for these estimators has already been developed (Baddeley and Turner (2005)).

As a final comment here, we have compared our proposed ABC-MCMC algorithm with the straightforward exchange algorithm of Murray et al. (2006), mentioned above, for a Strauss Gibbs point process. Without providing details, suppose we compute inefficiency factors (IF) for parameters, i.e., the ratio of the numerical variance of the estimate from the MCMC samples relative to that from hypothetical uncorrelated samples, using both model fitting approaches. We find that the IF for the exchange algorithm tend to be an order of magnitude greater than those from the ABC-MCMC algorithm. Also, the ABC-

MCMC algorithm allows simple parallelization, not possible for the exchange algorithm, so, computationally, it can be much faster.

3.3 Model Checking for Repulsive Point Processes

Model checking for spatial point processes has a relatively small literature. Approaches include (i) goodness of fit tests that can be used in conjunction with Poisson processes (e.g., Ripley (1988)) which emerge as special cases of Monte Carlo tests and (ii) residual analysis based on the Papangelou conditional intensity (see Baddeley et al. (2005)). For DPPs, likelihood ratio tests are available (see Lavancier et al. (2015)). However, such tests would not work to compare different types of repulsive point processes, i.e., DPPs with GPPs. For GPPs, without analytically available likelihoods, simulated comparison of the K or L functions under the fitted model with the observed are offered for model checking.

In the Bayesian framework, a cross validation approach based on training and test datasets can be applied for point processes with conditionally independent location densities, e.g., Poisson processes and Cox processes (Leininger and Gelfand (2016)). However, this approach is unavailable for repulsive point processes because holding out points will alter the geometric structure, hence the interaction structure, of the point pattern. As an alternative in our setting, we consider prior predictive Monte Carlo tests using the statistic $s_r(\mathbf{x})$ (5). For choices of r , we can implement the test for $s_r(\mathbf{y})$ together with the set $\{s_r(\mathbf{x}_u^*), u = 1, \dots, U\}$ where the \mathbf{x}_u^* 's are generated under the model with $\boldsymbol{\theta}_u \sim \pi(\boldsymbol{\theta})$.

For model comparison, we consider the ranked probability score (RPS, Gneiting and Raftery (2007) and Czado et al. (2009)) which assesses the performance of a predictive distribution relative to an observation, in our case to an observed count. Intuitively, a good model will provide a predictive distribution that is very concentrated around the observed count. For a set $B \subset D$, we calculate the RPS, via Monte Carlo integration, using posterior samples, as

$$RPS(F, N^{test}(B)) = \frac{1}{T} \sum_{t=1}^T |N^{(t)}(B) - N^{(test)}(B)| - \frac{1}{2T^2} \sum_{t=1}^T \sum_{t'=1}^T |N^{(t)}(B) - N^{(t')}(B)|. \quad (12)$$

Applying (12) to a collection of B_j uniformly drawn over D and summing over j gives a

model comparison criterion. Smaller values of the sum are preferred. We calculate RPS as in-sample comparison of predictive performance (see Leininger and Gelfand (2016)).

4 Simulation Studies

We consider simulation to address two points. In Section 4.1, we provide a proof of concept example to clarify how well we can recover model parameters using our ABC approach. In Section 4.2, we provide an example to illustrate model assessment.

4.1 Recovery of Model Parameters

First, we consider the strongly repulsive Strauss process on $D = [0, 1] \times [0, 1]$ as specified in (4),

$$\pi(\mathbf{x}) = \beta^{n(\mathbf{x})} \gamma^{s_R(\mathbf{x})} / c(\beta, \gamma), \quad s_R(\mathbf{x}) = \sum_{\{\xi, \eta\} \subseteq \mathbf{x}} \mathbf{1}[\|\xi - \eta\| \leq R] \quad (13)$$

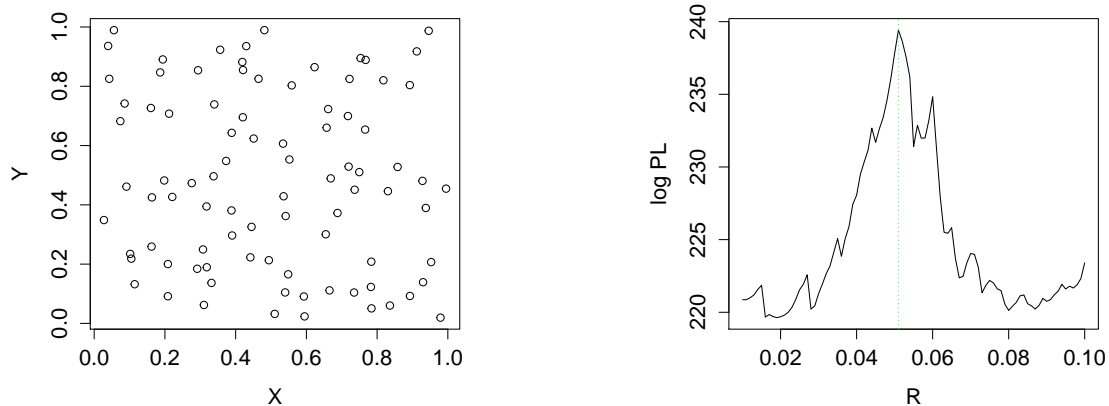
where $\beta = 200$, $\gamma = 0.1$, and $R = 0.05$. The realization is shown in the left panel of Figure 1. The number of points is $n = 88$. The prior specification is a uniform for β and γ : $\pi(\beta) = \mathcal{U}(50, 400)$, $\pi(\gamma) = \mathcal{U}(0, 1)$. The interaction radius R is estimated by the maximum pseudo-likelihood method. The estimated value is $\hat{R} = 0.051$ which is close to the true value (see the right panel of Figure 1). We use $\mathbf{T}_{suff} \equiv (\log n(\mathbf{x}), \sqrt{K_{\hat{R}}(\mathbf{x})})$ as our summary statistic. In the pilot run, we generate $L=10,000$ sets of $(\beta, \gamma, \mathbf{x})$ from $\pi(\mathbf{x}|\beta, \gamma) \times \pi(\beta)\pi(\gamma)$ under R fixed at \hat{R} , we calculate regression coefficients $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$, and we decide the truncation level ϵ based on the p_* % estimated percentiles of $\{\Psi(\hat{\boldsymbol{\theta}}_\ell, \hat{\boldsymbol{\theta}}_{obs})\}_{\ell=1}^L$. We set $\bar{\boldsymbol{\theta}} = \hat{\mathbf{a}}$ as the initial value and preserve 1,000 samples as posterior draws.

Table 1 shows the estimation results. The posterior for n is given though it is not a model parameter. We show three different acceptance levels. $\mathbf{T}_{suff}, \epsilon = 0$ provides the exact posterior distribution (due to sufficiency, and it can be observed due to discreteness). With $\mathbf{T}_{suff}, p_* = 1\%$ and $\mathbf{T}_{suff}, p_* = 5\%$, although the variance of estimated parameters increases slightly, the posterior means are well estimated.

As second example, we consider a DPP with a Gaussian covariance kernel (DPP-G),

$$C_G(x, y) = \tau \exp\left(-\frac{\|x - y\|^2}{\sigma^2}\right), \quad \tau = 100, \quad \sigma = 0.05. \quad (14)$$

Figure 1: Simulated realization (left) and profile pseudo likelihood (right) of a Strauss process



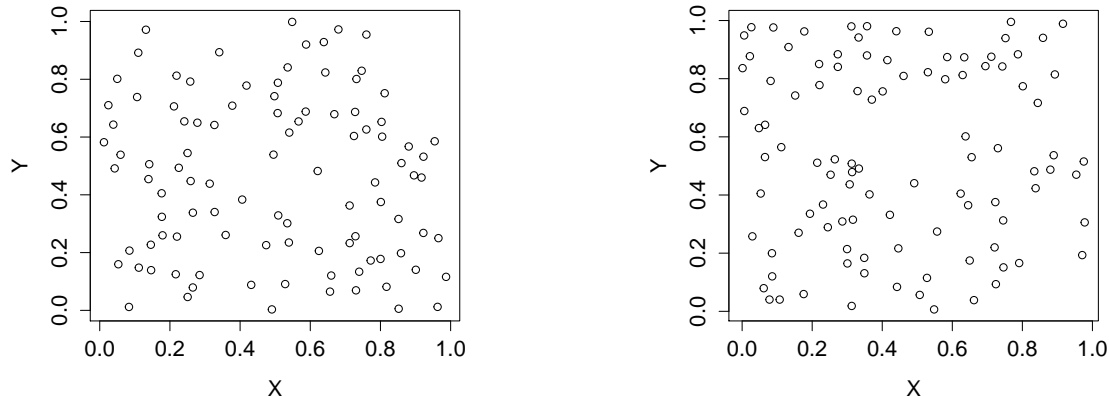
The realization is shown in the left panel of Figure 2 with the random $n = 100$. A uniform distribution for both parameters is assumed: $\pi(\tau) = \mathcal{U}(50, 200)$ (an interval which includes $\hat{\tau} = n/|D| = 100$) and $\pi(\sigma) = \mathcal{U}(0.001, \sigma_{max})$, where $\sigma_{max} = 1/\sqrt{\pi\tau} \approx 0.0564$ (required for the existence of a DPP with a Gaussian kernel). For DPP models, $\mathbf{T}_{func}(\mathbf{x}) = (\log n(\mathbf{x}), \sqrt{K_{r_1}(\mathbf{x})}, \dots, \sqrt{K_{r_M}(\mathbf{x})})$. We consider $M = 20$ and 10 equally spaced values over $[0, 0.1]$, i.e., $(r_1 = 0.005, r_2 = 0.01, \dots, r_{20} = 0.1)$ and $(r_1 = 0.01, r_2 = 0.02, \dots, r_{10} = 0.1)$. We implement a pilot run and then our ABC algorithm with the same number of posterior samples as in the first example. Table 2 illustrates that the true values of parameters are recovered well. The estimation results are insensitive to the choice of M . The posterior variance of the parameters increases slightly when the tolerable level based on p_* increases but the posterior means are well estimated.

As third example, we consider a DPP with power exponential spectral density (DPP-PE),

$$\varphi(x) = \tau \frac{\alpha^2}{\pi\Gamma(2/\nu + 1)} \exp(-\|\alpha x\|^\nu), \quad \tau = 100, \quad \alpha = 0.1, \quad \nu = 10(\text{fixed}) \quad (15)$$

This kernel can capture stronger repulsiveness than a DPP with a Gaussian kernel (DPP-G). Though the parameters in this kernel are not interpretable, we can still investigate whether or not they are recovered. The random number of points is $n = 101$. A uniform distribution for τ and α is assumed: $\pi(\tau) = \mathcal{U}(50, 200)$ and $\pi(\alpha) = \mathcal{U}(0.001, \alpha_{max})$ where

Figure 2: Simulated realization of a DPP with Gaussian kernel (left) and a DPP with power exponential spectral (right)



$\alpha_{max} = \sqrt{\Gamma(2/\nu + 1)\pi/\tau} \approx 0.1698$. Again, we implement a pilot run and then our ABC algorithm with the same number of posterior samples. Table 2 demonstrates that the true values of parameters are recovered well.

4.2 Model Assessment

As in the discussion in Section 3.3, we examine model adequacy through Monte Carlo tests. We consider two different point pattern sizes: (A) $n \approx 100$ and (B) $n \approx 500$. Four true datasets are generated for each pattern size. For (A): (A.i) HPP ($\lambda = 100$), (A.ii) Strauss process ($\beta = 250, \gamma = 0.05, R = 0.05$), (A.iii) DPP-G ($\tau = 100, \sigma = \sigma_{max}$), (A.iv) DPP-PE ($\tau = 100, \alpha = \alpha_{max}, \nu = 10$). For (B): (B.i) HPP ($\lambda = 500$), (B.ii) Strauss process ($\beta = 1000, \gamma = 0.05, R = 0.02$), (B.iii) DPP-G ($\tau = 500, \sigma = \sigma_{max}$) and (B.iv) DPP-PE ($\tau = 500, \alpha = \alpha_{max}, \nu = 10$). The DPP-PE with $\nu = 10$ and $\alpha = \alpha_{max}$ provides stronger repulsion than the DPP-G model with $\sigma = \sigma_{max}$ (see, Lavancier et al. (2015)). All settings are expected to have 100 points for (A) and 500 points for (B), respectively so we focus on comparing the second order statistics $s_r(\mathbf{x})$.

For each model, the prior specifications are as follows. Let $\mathcal{G}(\cdot)$, $\mathcal{B}(\cdot)$ and $\mathcal{U}(\cdot)$ denote Gamma, Beta and uniform distributions respectively. Then, for (A), HPP: $\pi(\lambda) = \mathcal{G}(200, 2)$ ($E[\lambda] = 100$, $\text{Var}[\lambda] = 50$), Strauss: $\pi(\beta) = \mathcal{U}(75, 400)$ $\pi(\gamma) = \mathcal{B}(1, 6)$, DPP-G: $\pi(\tau) =$

Table 1: Estimation results for the Strauss process

Strauss	True	Mean	Stdev	95% Int
$\mathbf{T}_{suff} \epsilon = 0$				
β	200	180.9	23.65	[139.3, 233.5]
γ	0.1	0.165	0.062	[0.066, 0.309]
n	88	88	0	[88, 88]
$\mathbf{T}_{suff} p_* = 1\%$				
β	200	180.7	24.53	[135.0, 230.6]
γ	0.1	0.167	0.071	[0.057, 0.336]
n	88	88	1.593	[86, 91]
$\mathbf{T}_{suff} p_* = 5\%$				
β	200	181.5	31.08	[127.1, 247.0]
γ	0.1	0.157	0.091	[0.020, 0.361]
n	88	88	5.705	[77, 97]

$\mathcal{G}(200, 2)$ $\pi(\sigma/\sigma_{max}) = \mathcal{B}(6, 1)$, DPP-PE: $\pi(\tau) = \mathcal{G}(200, 2)$ $\pi(\alpha/\alpha_{max}) = \mathcal{B}(6, 1)$. The beta priors for γ , α/α_{max} and σ/σ_{max} imply moderate to strong repulsion within each model because each model shows stronger repulsiveness when γ is small and the α/α_{max} and σ/σ_{max} are large.

For (B) HPP: $\pi(\lambda) = \mathcal{G}(1000, 2)$ ($E[\lambda] = 500$, $\text{Var}[\lambda] = 250$), Strauss: $\pi(\beta) = \mathcal{U}(350, 1200)$ $\pi(\gamma) = \mathcal{B}(1, 6)$, DPP-G: $\pi(\tau) = \mathcal{G}(1000, 2)$ $\pi(\sigma/\sigma_{max}) = \mathcal{B}(6, 1)$ where $\sigma_{max} = 1/\sqrt{\pi\tau}$, DPP-PE: $\pi(\tau) = \mathcal{G}(1000, 2)$ $\pi(\alpha/\alpha_{max}) = \mathcal{B}(6, 1)$.

Table 3 shows estimated p -values for the Monte Carlo tests associated with the observed $s_r(\mathbf{y})$ by simulating 999 point patterns $\{\mathbf{y}^*\}$ from the prior predictive distribution of each model. For $n \approx 100$, models are assessed at the radii $r = (0.03, 0.05, 0.07, 0.09)$; for $n \approx 500$ models are assessed at the radii $r = (0.01, 0.02, 0.03, 0.04)$. For $n \approx 100$ and $n \approx 500$ we see similar results. When the true model is the HPP or the Strauss process, the other models are sometimes formally criticized but, regardless, show smaller p -values. When the DPP-G is the true model, only the HPP is criticized. When DPP-PE is true, DPP-G and HPP are criticized but Strauss is not. This result is not surprising because, though the

Table 2: Estimation results for the DPP with Gaussian kernel (left) and with power exponential spectral density (right) for each number of evaluation grids and percentiles.

Gauss (M, p_* %)	True	Mean	Stdev	95% Int	PE	True	Mean	Stdev	95% Int
(10, 1%)									
τ	100	100.4	5.863	[89.84, 112.5]	τ	100	100.9	6.011	[88.47, 112.6]
σ	0.05	0.047	0.006	[0.030, 0.057]	α	0.1	0.113	0.022	[0.069, 0.155]
n	100	100	3.198	[94, 106]	n	101	101	2.99	[96, 107]
(10, 5%)									
τ	100	100.0	8.205	[85.26, 117.1]	τ	100	102.9	7.899	[87.93, 117.3]
σ	0.05	0.045	0.008	[0.026, 0.057]	α	0.1	0.112	0.025	[0.060, 0.160]
n	100	100	7.026	[88, 113]	n	101	102	6.307	[90, 113]
(20, 1%)									
τ	100	99.28	6.277	[86.30, 110.7]	τ	100	102.3	6.568	[88.93, 114.8]
σ	0.05	0.048	0.006	[0.032, 0.057]	α	0.1	0.110	0.22	[0.061, 0.151]
n	100	100	3.140	[94, 105]	n	101	101	2.922	[96, 106]
(20, 5%)									
τ	100	99.06	7.434	[84.32, 113.6]	τ	100	101.4	7.490	[86.02, 116.0]
σ	0.05	0.045	0.007	[0.028, 0.057]	α	0.1	0.110	0.025	[0.057, 0.157]
n	100	99	6.596	[88, 111]	n	101	101	6.313	[90, 113]

DPP-PE presents stronger repulsiveness than the DPP-G, the Strauss process shows very strong repulsiveness. As expected, the results emerge more strongly for the larger point patterns.

We also investigate in-sample RPS to compare the model fitting. We set $J = 1000$ and sample squares B_j uniformly over D . Each B_j is a square of size $q|D|$ with $q \in (0, 0.1)$. We calculate RPS for each B_j and average over $j = 1, \dots, J$. Figure 3 shows the relative RPS, i.e., the ratio of RPS for a particular model relative to the RPS for the true model. For $n \approx 500$, though the ratios are close to 1.00, the true model is generally preferred. However, when the DPP-G is true, the difference between DPP-G and DPP-PE is small.

For $n \approx 100$, the comparison is a bit less successful. When the HPP or the DPP-PE are true, these models are preferred by the in-sample RPS. However, when the Strauss or the DPP-G are true, they don't show the smallest RPS, although the difference is small. Moreover, even for $n \approx 100$, the HPP can be distinguished from other repulsive point processes even with moderate repulsiveness. Altogether, we conclude that a larger number of points will be required to distinguish the true repulsive point process from the other repulsive point processes.

5 Real Data Application

We implement our approach for a dataset of tree locations in Duke Forest, comprising 68 species, a total of 13,655 individuals with diameter at breast height (dbh) and location recorded for each. We aggregate the species and remove trees under 40 dbh from the dataset because only older trees exhibit inhibition/repulsion. The left panel in Figure 4 shows the locations of trees whose dbh is larger than 40 over a selected rectangle (rescaled to the unit square for fitting) in the Blackwood region of Duke Forest. The resulting number of points is 89.

For the Strauss model, we include both an interaction radius $R \approx 0.053$ and a hardcore radius which are chosen from profile pseudo likelihoods (see, the middle panel in Figure 4)². We also investigated the cases $R = 0.02$ and $R = 0.035$. The prior specifications for the Strauss process are: $\beta \sim \mathcal{U}(50, 350)$, $\gamma \sim \mathcal{U}(0, 1)$. We fix $p_* = 1\%$ percentiles to determine ϵ for Strauss models. We also consider a DPP-G and a DPP-PE with $\tau \sim \mathcal{U}(50, 200)$ and $\sigma \sim \mathcal{U}(0.001, \sigma_{max})$ where $\sigma_{max} = 1/\sqrt{\pi\tau}$ for the DPP-G and $\tau \sim \mathcal{U}(50, 200)$ and $\alpha \sim \mathcal{U}(0.001, \alpha_{max})$ where $\alpha_{max} = \sqrt{\Gamma(2/\nu + 1)\pi/\tau}$ for the DPP-PE. Through the pilot run, we fix $M = 10$ and $p_* = 1\%$ percentiles to determine ϵ for DPP models. We preserve 1,000 samples for each model. Table 4 shows the estimation results for the HPP, the Strauss, the DPP-G and the DPP-PE models. The estimated value of γ for the Strauss process with $R = 0.053$ reveals a moderate level of repulsion.

As above, we consider prior predictive checks using Monte Carlo tests employing $s_r(\mathbf{y})$

²In the simulation examples, we considered only an interaction radius. However, with real data, often we find hardcore repulsion with a very small radius or moderate repulsion with a larger radius.

Table 3: Model Adequacy: estimated p -values for simulated datasets, $n \approx 100$ (left) and $n \approx 500$ (right). Bold denotes significance at $p \leq .1$.

$n \approx 100 \backslash r$	0.03	0.05	0.07	0.09	$n \approx 500 \backslash r$	0.01	0.02	0.03	0.04
HPP					HPP				
HPP	0.669	0.588	0.592	0.578	HPP	0.404	0.638	0.596	0.636
Strauss	0.012	0.011	0.077	0.137	Strauss	0.017	0.009	0.086	0.147
DPP-G	0.005	0.035	0.130	0.215	DPP-G	0.001	0.001	0.013	0.028
DPP-PE	0.067	0.149	0.217	0.257	DPP-PE	0.001	0.001	0.001	0.034
Strauss					Strauss				
HPP	0.001	0.001	0.092	0.268	HPP	0.001	0.001	0.001	0.038
Strauss	0.481	0.450	0.697	0.719	Strauss	0.409	0.382	0.617	0.638
DPP-G	0.105	0.001	0.345	0.469	DPP-G	0.114	0.001	0.239	0.369
DPP-PE	0.127	0.001	0.333	0.452	DPP-PE	0.413	0.001	0.745	0.735
DPP-G					DPP-G				
HPP	0.005	0.020	0.108	0.236	HPP	0.001	0.001	0.001	0.034
Strauss	0.386	0.101	0.285	0.303	Strauss	0.409	0.768	0.380	0.492
DPP-G	0.245	0.274	0.364	0.441	DPP-G	0.328	0.199	0.260	0.355
DPP-PE	0.224	0.275	0.357	0.421	DPP-PE	0.709	0.788	0.764	0.721
DPP-PE					DPP-PE				
HPP	0.001	0.003	0.017	0.090	HPP	0.001	0.001	0.001	0.001
Strauss	0.481	0.833	0.502	0.547	Strauss	0.409	0.768	0.380	0.492
DPP-G	0.105	0.075	0.104	0.201	DPP-G	0.114	0.005	0.003	0.036
DPP-PE	0.127	0.122	0.131	0.209	DPP-PE	0.413	0.160	0.144	0.199

Figure 3: In sample relative RPS: $n \approx 100$ (left) and $n \approx 500$ (right)

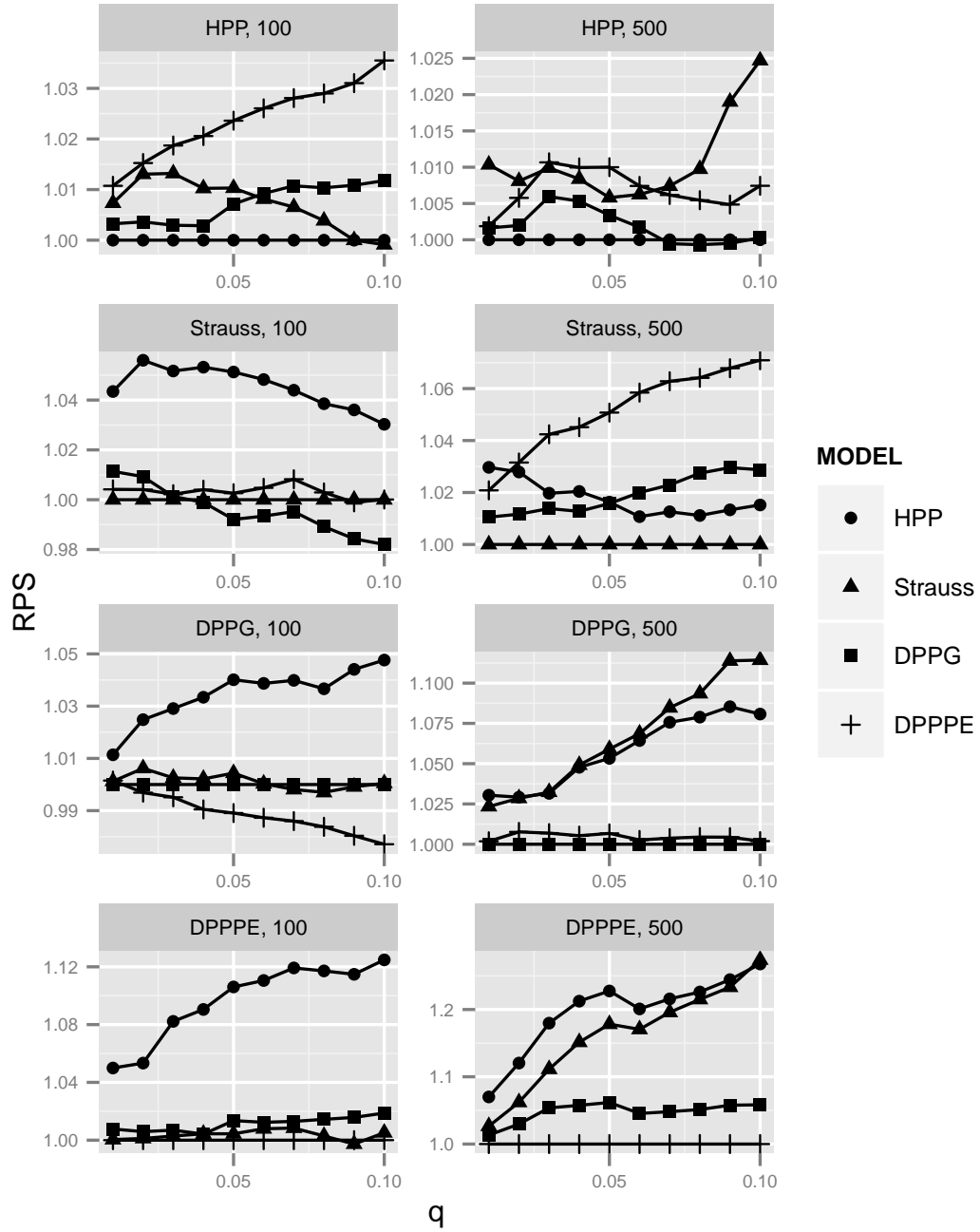


Figure 4: Plot of Duke Forest dataset (left), profile pseudolikelihood for the Strauss process (middle) and $\hat{L}(r) - r$ for the focused region (right)

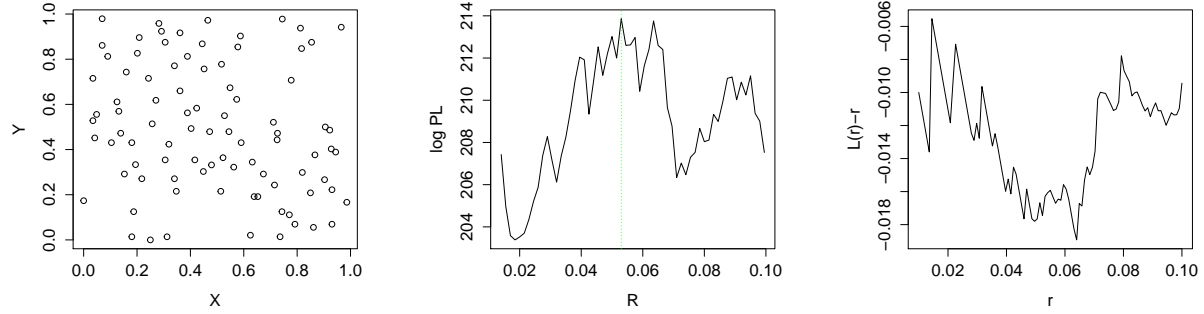


Table 4: Estimation results for the real data

	Mean	Stdev	95% Int		Mean	Stdev	95% Int
HPP				Strauss			
				$R = 0.02$			
β	89.75	5.464	[79.73, 100.4]	β	96.55	10.34	[77.11, 118.3]
n	90	10.79	[70, 112]	γ	0.452	0.243	[0.065, 0.937]
				n	89	3.079	[84, 94]
DPP-G				Strauss			
				$R = 0.035$			
τ	88.96	6.446	[76.94, 101.8]	β	111.6	13.25	[86.99, 138.5]
σ	0.048	0.009	[0.025, 0.060]	γ	0.467	0.185	[0.169, 0.866]
n	89	4.287	[82, 97]	n	90	2.446	[85, 93]
DPP-PE				Strauss			
				$R = 0.053$			
τ	89.47	4.936	[79.16, 98.71]	β	149.0	23.43	[110.3, 200.3]
α	0.144	0.025	[0.084, 0.179]	γ	0.452	0.140	[0.202, 0.744]
n	90	3.058	[84, 95]	n	90	1.109	[88, 92]

Table 5: Model Adequacy: estimated p -values for the real dataset. Bold denotes significance at $p \leq .1$

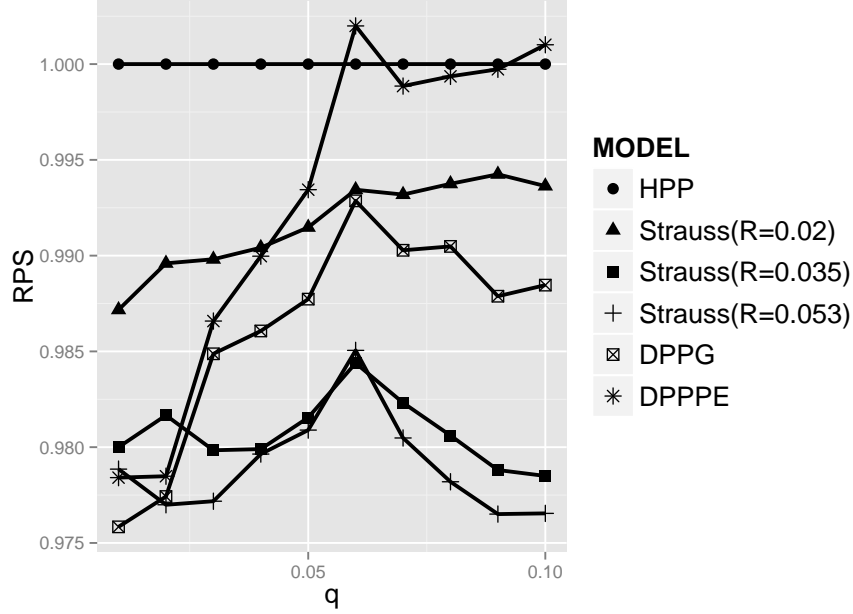
Model\(r	0.03	0.05	0.07	0.09
HPP	0.0428	0.0130	0.1002	0.1662
Strauss ($R = 0.02$)	0.0572	0.0528	0.0919	0.1040
Strauss ($R = 0.035$)	0.2373	0.0978	0.1292	0.1392
Strauss ($R = 0.053$)	0.3574	0.3923	0.3340	0.3042
DPP-G	0.1894	0.1311	0.1978	0.2097
DPP-PE	0.2828	0.1801	0.2274	0.2305

for comparing second order properties. Table 5 presents the estimated p -values for the observed $s_r(\mathbf{y})$ for each model and radius. Including prior specification, Table 5 reveals that the HPP is criticized for small radii. The Strauss model with small radii ($R = 0.02$ and $R = 0.035$) are also criticized. There is evidence in support of repulsion with moderate radius. We also calculate in-sample RPS. Figure 5 shows the relative RPS to the RPS of the HPP. Again, the ratios are near 1.00. However, despite the somewhat small size of the point pattern, the figure indicates preference for the repulsive point processes, most strongly for the Strauss model with $R = 0.053$.

6 Summary and future work

Bayesian inference for DPPs and GPPs is not unified because of challenging model fitting problems specific to each type of process model. Here, we have proposed a unifying approach for model fitting using ABC. It is attractive because simulation from the models is easier than evaluation of the exact likelihood and because informative summary statistics (first and second moments) are available. We also offered model assessment strategies for repulsive point processes using Monte Carlo tests and in-sample RPS. Simulation studies illustrate that true models can be recovered but that it may be difficult to criticize the true repulsive point process in favor of other repulsive specifications when the number of points is small.

Figure 5: In-sample relative RPS for Duke forest data



Future work will examine nonhomogeneous repulsive point processes. Because the second order functional summary statistics are available, e.g., the nonhomogeneous versions of the K function (Baddeley et al. (2000)), our ABC approach can be directly extended. Another challenge is to consider very large datasets. The main computational cost of our algorithm is simulation of \mathbf{x} for each iteration of the ABC-MCMC. One promising solution is to run multiple MCMC chains in parallel (Wilkinson (2005)). Through the pilot run, we can obtain the rough approximate posterior mean of the parameters $\bar{\hat{\theta}} = \hat{\mathbf{a}}$. Starting the algorithm with this initial value results in no need for a burn-in period. In this regard, other types of ABC algorithms can be considered, for example, sequential versions of ABC (ABC-SMC) Beaumont et al. (2009), Toni et al. (2009) and Marin et al. (2012) which might be more suitable for parallelization.

Acknowledgements

The work of the first author was supported in part by the Nakajima Foundation. The authors thank James Clark for providing the Duke Forest dataset.

SUPPLEMENTARY MATERIAL

R-code R-code is attached for Section 4 and 5

Duke Forest data set: Data set used in Section 5

References

- Affandi, R. H., E. B. Fox, R. P. Adams, and B. Taskar (2014). Learning the parameters of determinantal point process kernels. In *Thirty-First International Conference on Machine Learning (ICML)*.
- Affandi, R. H., E. B. Fox, and B. Taskar (2013). Approximate inference in continuous determinantal processes. In *Neural Information Processing Systems (NIPS)*.
- Baddeley, A., J. Møller, and R. Waagepetersen (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica* 54, 329–350.
- Baddeley, A. and R. Turner (2000). Practical maximum pseudolikelihood for spatial point patterns. *Aust N Z J Stat* 42, 283–315.
- Baddeley, A. and R. Turner (2005). **spatstat**: an R package for analyzing spatial point patterns. *Journal of Statistical Software* 12, 1–42.
- Baddeley, A., R. Turner, J. Møller, and M. Hazelton (2005). Residual analysis for spatial point processes. *Journal of the Royal Statistical Society, Series B* 67, 617–666.
- Beaumont, M. A., J. M. Cornuet, J. M. Marin, and C. P. Robert (2009). Adaptive approximate Bayesian computation. *Biometrika* 96, 983–990.
- Beaumont, M. A., W. Zhang, and D. J. Balding (2002). Approximate Bayesian computation in population genetics. *Genetics* 162, 2025–2035.
- Berthelsen, K. K. and J. Møller (2002). A primer on perfect simulation for spatial point processes. *Bulletin of the Brazilian Mathematical Society* 33, 351–367.

- Berthelsen, K. K. and J. Møller (2003). Likelihood and non-parametric Bayesian MCMC inference for spatial point processes based on perfect simulation and path sampling. *Scandinavian Journal of Statistics* 30, 549–564.
- Berthelsen, K. K. and J. Møller (2008). Non-parametric Bayesian inference for inhomogeneous Markov point processes. *Aust N Z J Stat* 50, 257–272.
- Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B* 48, 259–302.
- Besag, J. E. (1975). Statistical analysis of non-lattice data. *The Statistician* 24, 179–195.
- Besag, J. E. (1977). Discussion to the paper: Modelling spatial patterns. by B. D. Ripley. *Journal of the Royal Statistical Society, Series B* 39, 172–212.
- Blum, M. G. B. and V. C. Tran (2010). HIV with contact tracing: a case study in approximate Bayesian computation. *Biostatistics* 11, 644–660.
- Czado, C., T. Gneiting, and L. Held (2009). Predictive model assessment for count data. *Biometrics* 65, 1254–1261.
- Diggle, P. (2003). *Statistical Analysis of Spatial Point Patterns*, 2nd ed. Hodder Arnold.
- Drovandi, C. C. and A. N. Pettitt (2011). Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. *Biometrics* 67, 225–233.
- Erhardt, R. J. and R. L. Smith (2012). Approximate Bayesian computing for spatial extremes. *Computational Statistics and Data Analysis* 56, 1468–1481.
- Fearnhead, P. and D. Prangle (2012). Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. *Journal of the Royal Statistical Society, Series B* 74, 419–474.
- Friedman, J., T. Hastie, and R. Tibshirani (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33, 1–22.
- Georgii, H. O. (1988). *Gibbs measures and phase transition*. Walter de Gruyter.

- Geyer, C. J. and J. Møller (1994). Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics* 21, 359–373.
- Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 102, 359–378.
- Goldstein, J., M. Haran, I. Simeonov, J. Fricks, and F. Chiaromonte (2015). An attraction-repulsion point process model for respiratory syncytial virus infections. *Biometrics* 71, 376–385.
- Guan, Y. (2006). A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association* 101, 1502–1512.
- Hough, J. B., M. Krishnapur, Y. Peres, and B. Virag (2006). Determinantal processes and independence. *Probability Surveys* 3, 206–229.
- Huang, F. and Y. Ogata (1999). Improvements of the maximum pseudo-likelihood estimators in various spatial statistical models. *Journal of Computational and Graphical Statistics* 8, 510–530.
- Illian, J., A. Penttinen, H. Stoyan, and D. Stoyan (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. Wiley.
- Jensen, J. L. and J. Møller (1991). Pseudolikelihood for exponential family models of spatial point processes. *Annals of Applied Probability* 1, 445–461.
- Kelly, F. P. and B. D. Ripley (1976). A note on Strauss’ model for clustering. *Biometrika* 63, 357–360.
- Kendall, W. S. and J. Møller (2000). Perfect simulation using dominating processes on ordered spaces, with application to locally stable point processes. *Advances in Applied Probability* 32, 844–865.
- King, R., J. B. Illian, S. E. King, G. F. Nightingale, and D. K. Hendrichsen (2012). A Bayesian approach to fitting Gibbs processes with temporal random effects. *Journal of Agricultural Biological and Environmental Statistics* 17, 601–622.

- Kulesza, A. and B. Taskar (2012). Determinantal point processes for machine learning. *Foundations and Trends in Machine Learning* 5, 2–3.
- Lavancier, F., J. Møller, and E. Rubak (2015). Determinantal point processes models and statistical inference. *Journal of the Royal Statistical Society, Series B* 77, 853–877.
- Leininger, T. J. and A. E. Gelfand (2016). Bayesian inference and model assessment for spatial point patterns using posterior predictive samples. *Bayesian Analysis*. forthcoming.
- Liang, F. (2009). A double Metropolis-Hastings sampler for spatial models with intractable normalizing constants. *Journal of Statistical Computation and Simulation* 80, 1007–1022.
- Macchi, O. (1975). The coincidence approach to stochastic point processes. *Advances in Applied Probability* 7, 83–122.
- Marin, J.-M., P. Pudlo, C. P. Robert, and R. J. Ryder (2012). Approximate Bayesian computational methods. *Statistics and Computing* 22, 1167–1180.
- Marjoram, P., J. Molitor, V. Plagnol, and S. Tavaré (2003). Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences* 100, 15324–15328.
- Mase, S. (1995). Consistency of the maximum pseudo-likelihood estimator of continuous state space Gibbsian processes. *Annals of Applied Probability* 5, 603–612.
- Mattfeldt, T., S. Eckel, F. Fleischer, and V. Schmidt (2007). Statistical modelling of the geometry of planar sections of prostatic capillaries on the basis of stationary Strauss hard-core processes. *Journal of Microscopy* 228, 272–81.
- Møller, J., A. N. Pettit, R. Reeves, and K. K. Berthelsen (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika* 93, 451–458.
- Møller, J. and R. Waagepetersen (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/RC.

- Murray, I., Z. Ghahramani, and D. J. C. MacKay (2006). MCMC for doubly-intractable distributions. In *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence (UAI)*. AUAI Press.
- Picchini, U. (2013). Inference for SDE models via approximate Bayesian computation. *Journal of Computational and Graphical Statistics* 23, 1080–1100.
- Pritchard, J., M. Seielstad, A. Perez-Lezaun, and M. Feldman (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution* 16, 1791–1798.
- Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability* 13, 255–266.
- Ripley, B. D. (1977). Modelling spatial patterns. *Journal of the Royal Statistical Society, Series B* 39, 172–212.
- Ripley, B. D. (1988). *Statistical inference for spatial processes*. Cambridge University Press.
- Scardicchio, A., C. Zachary, and S. Torquato (2009). Statistical properties of determinantal point processes in high-dimensional Euclidean spaces. *Physical Review E* 79.
- Sisson, S. A. and Y. Fan (2011). Likelihood-free Markov chain Monte Carlo. In *Handbook of Markov Chain Monte Carlo*, pp. 313–338. Chapman and Hall/CRC.
- Soubeyrand, S., F. Carpentier, F. Guiton, and E. K. Klein (2013). Approximate Bayesian computation with functional statistics. *Statistical Applications in Genetics and Molecular Biology* 12, 17–37.
- Stoyan, D. and A. Penttinen (2000). Recent applications of point process methods in forestry statistics. *Statistical Science* 15, 61–78.
- Strauss, D. J. (1975). A model for clustering. *Biometrika* 62, 467–475.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B* 58, 267–288.

- Toni, T., D. Welch, N. Strelkowa, A. Ipsen, and M. Stumpf (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamic systems. *Journal of the Royal Society Interface* 6, 187–202.
- van Lieshout, M. N. M. (2000). *Markov point processes and their applications*. Imperial College Press.
- Wilkinson, D. J. (2005). Parallel Bayesian computation. In *Handbook of Parallel Computing and Statistics*, pp. 481–512. Marcel Dekker/CRC.